Arch Virol (2003) 148: 1925–1941
DOI 10.1007/s00705-003-0147-0

Archives of Virology
Printed in Austria

Variability and transmission by *Aphis glycines* of North American and Asian *Soybean mosaic virus* isolates

L. L. Domier[1,2], **I. J. Latorre**[1], **T. A. Steinlage**[1], **N. McCoppin**[1,2], and **G. L. Hartman**[1,2]

[1]Department of Crop Sciences, University of Illinois at Urbana/Champaign, Urbana, Illinois, U.S.A.
[2]United States Department of Agriculture–Agricultural Research Service, University of Illinois at Urbana/Champaign, Urbana, Illinois, U.S.A.

Received March 20, 2002; accepted May 14, 2003
Published online July 29, 2003

Summary. The variability of North American and Asian strains and isolates of *Soybean mosaic virus* was investigated. First, polymerase chain reaction (PCR) products representing the coat protein (CP)-coding regions of 38 SMVs were analyzed for restriction fragment length polymorphisms (RFLP). Second, the nucleotide and predicted amino acid sequence variability of the P1-coding region of 18 SMVs and the helper component/protease (HC/Pro) and CP-coding regions of 25 SMVs were assessed. The CP nucleotide and predicted amino acid sequences were the most similar and predicted phylogenetic relationships similar to those obtained from RFLP analysis. Neither RFLP nor sequence analyses of the CP-coding regions grouped the SMVs by geographical origin. The P1 and HC/Pro sequences were more variable and separated the North American and Asian SMV isolates into two groups similar to previously reported differences in pathogenic diversity of the two sets of SMV isolates. The P1 region was the most informative of the three regions analyzed. To assess the biological relevance of the sequence differences in the HC/Pro and CP coding regions, the transmissibility of 14 SMV isolates by *Aphis glycines* was tested. All field isolates of SMV were transmitted efficiently by *A. glycines*, but the laboratory isolates analyzed were transmitted poorly. The amino acid sequences from most, but not all, of the poorly transmitted isolates contained mutations in the aphid transmission-associated DAG and/or KLSC amino acid sequence motifs of CP and HC/Pro, respectively.

Introduction

Soybean mosaic virus (SMV), a member of the genus *Potyvirus* in the family *Potyviridae* [4], causes major diseases of soybean (*Glycine max* [L.] Merr) that reduce soybean yields from 8 to 35% and occurs in virtually all soybean production areas of the world [20]. SMV particles are flexuous filaments, approximately 750 nm long and 15 to 18 nm in diameter that are composed of a 9.6-kb single-stranded positive-sense RNA molecule associated with a 29.9-kDa coat protein (CP) [20]. The CP-coding and complete genomic nucleotide sequences have been determined for several SMV isolates [10, 15, 25, 26, 37].

SMV isolates can produce very different symptoms on soybean plants. Consequently, SMV isolates were first discriminated based on symptoms produced on a differential set of soybean lines, which defined SMV strains G1 through G7 [8]. SMV strains C14 [36] and G7a [9] were added subsequently to describe isolates that differed from the type strains in symptoms on the differential set of soybean lines. Strains G8 through G11 [67] were added to describe the wider diversity in pathogenicity of isolates present in China. At least four soybean genes have been reported to confer resistance to different strains of SMV [7, 28, 36]. SMV isolates also have been grouped on the basis of vector transmission [38], antigenic properties [18, 19], tryptic peptide profiles [24], and restriction enzyme polymorphism (RFLP) analysis of reverse transcriptase-polymerase chain reaction (RT-PCR) products [44].

SMV is transmitted efficiently through seed and by aphids of at least 32 species in a nonpersistent manner [20]. Prior to the identification of *Aphis glycines* in 2000, no aphid species that colonized soybean were known in North America. The introduction of *A. glycines*, which transmits SMV and several other soybean-infecting viruses [11, 21, 64], has the potential to greatly increase the incidence of viruses that are persistently or nonpersistently transmitted by aphids. Two primary determinants of aphid transmission of potyviruses, the helper component protease (HC/Pro) and CP, have been identified and their interactions with aphid stylets and each other has been studied extensively [46]. HC/Pro functions as a bridge between aphid stylets and virus particles. The cysteine-rich amino acid terminus of HC/Pro interacts with aphid stylets through a conserved amino acid motif (KITC) [6] and with the DAG motif in the CP through a second conserved amino acid sequence (PTK). The DAG amino acid motif is exposed on the surface of virions and has been associated with aphid transmissibility of several potyviruses [2, 45, 48]. While some variability is tolerated in the KITC and DAG motifs, the PTK motif is retained in all aphid transmissible potyviruses [46]. Unrelated to its functions in aphid transmission, HC/Pro also plays a major role in suppression of post-transcriptional gene silencing in infected plants [56].

Within the family *Potyviridae*, nucleotide and amino acid sequences of CP genes have been used to assess phylogenetic relationships between species [3, 65]. However, because CP sequences are highly conserved among members of the *Potyviridae* and represent only about 8% of the genome, CP regions often do not contain sufficient diversity to differentiate closely related isolates of the same virus species. Therefore, other regions of potyvirus genomes have been evaluated phylogenetically, including noncoding regions [41, 60] and regions encoding, the amino-terminal region of the polyprotein (P1) [1, 34], HC/Pro [1, 60], nuclear inclusion b [1, 62], and cytoplasmic inclusion [33]. HC/Pro proteins are multifunctional with variable amino termini and well conserved carboxyl termini [48]. P1 proteins possess protease activities [61] and are the most variable and phylogenetically informative region of potyvirus genomes [50].

In this study, the genetic relatedness and molecular variability of SMV isolates from North America and Asia were assessed by examining nucleotide and predicted amino acid sequence data from three genomic regions, P1, HC/Pro, and CP. In addition, the transmissibility by *A. glycines* of a subset of the isolates was assessed.

Methods

Virus isolates and strains

A total of 38 SMV strains and isolates were used in this study (Table 1). SMV strains G1, G2, G3, G6, and G7 were provided by Dr. J. Hill (Iowa State University, Ames, IA); G5 (PV-573) was obtained from American Type Culture Collection (Manassas, VA); and G7a and G7F were obtained from original cultures of Cho and Goodman [8], stored at the University of Illinois at Urbana-Champaign. The strain designations of G1, G2, G3, G5, G6, G7, G7a and G7F were not confirmed on differential hosts for these studies. Other previously uncharacterized North American SMV isolates were obtained from the University of Illinois South Farms, and Virginia (provided by Dr. S. A. Tolin, Virginia Polytechnic Institute and State University, Blacksburg, VA). Asian SMV isolates were obtained from soybean seed from central and northwestern China and Korea (provided by R. L. Nelson, USDA Soybean Germplasm Collection, Urbana, IL). Asian SMV isolates were named for the province

Table 1. Soybean mosaic virus isolates

Designation	Origin	Source	GenBank
G1	Illinois/USA	John Hill	AH008451
G2	Illinois/USA	John Hill	S42280
G3	Illinois/USA	John Hill	AH008452
G5	Illinois/USA	ATCC	AH008453
G6	Illinois/USA	John Hill	AH008454
G7	Illinois/USA	John Hill	AF241739
G7a	Illinois/USA	ATCC	AH008455
G7F	Illinois/USA	UIUC	AH008456
Ch1	China	germplasm	
Ch2	China	germplasm	
Ch3	China	germplasm	
Ch4	China	germplasm	
ChF	Fujian/China	germplasm	
ChGs1	Gansu/China	germplasm	AH008404
ChGs2	Gansu/China	germplasm	AH008402
ChGs3	Gansu/China	germplasm	AH008447
ChGs4	Gansu/China	germplasm	
ChGs5	Gansu/China	germplasm	
ChGs6	Gansu/China	germplasm	
ChGs7	Gansu/China	germplasm	

(*continued*)

Table 1 (*continued*)

Designation	Origin	Source	GenBank
ChH1	Hebei/China	germplasm	AH008448
ChH2	Hebei/China	germplasm	AH008403
ChH3	Hebei/China	germplasm	
ChJ	Jiangsu/China	germplasm	AH008449
ChN	Ningxia/China	germplasm	
ChS1	Shanxi/China	germplasm	AH008450
ChS2	Shanxi/China	germplasm	
ChS3	Shanxi/China	germplasm	
ChS4	Shanxi/China	germplasm	
ChS5	Shanxi/China	germplasm	
IL1	Illinois/USA	field collection	AH008458
IL2	Illinois/USA	field collection	
IL3	Illinois/USA	field collection	
IL4	Illinois/USA	field collection	
IL5	Illinois/USA	field collection	
413	Illinois/USA	field collection	AH012606
452	Illinois/USA	field collection	AH012607
746	Illinois/USA	field collection	AH012608
770	Illinois/USA	field collection	AH012609
1071	Illinois/USA	field collection	AH012604
1083	Illinois/USA	field collection	AH012605
SK	South Korea	germplasm	
VA1	Virginia/USA	Sue Tolin	AH008457
VA2	Virginia/USA	Sue Tolin	AH008459
VA3	Virginia/USA	Sue Tolin	
VA4	Virginia/USA	Sue Tolin	

from which the soybean seed originated. All laboratory isolates were maintained by mechanical inoculation in *G. max* var. Williams 82.

Enzyme-linked immunosorbent assay

All SMV infections were confirmed by double antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA) in Immulon U microtiter plates (Dynex Technologies, Chantilly, VA) using SMV polyclonal antibodies (Agdia, Elkhart, IN) according to the supplier's recommendations. Optical densities were determined at 405 nm with a Dynex MRX automated plate reader. Samples were considered infected when the absorbance was greater than twice the mean absorbance value of sap from greenhouse-grown healthy soybean leaves. Extracts from greenhouse-grown soybean leaves infected with SMV-G7a were used as positive controls.

Reverse transcription polymerase chain reaction

Total RNA was extracted from approximately 0.3 grams of soybean tissue using Trizol regent (Invitrogen, Carlsbad, CA), precipitated, and resuspended in 100 μl 10 mM Tris-HCl (pH 7.4), 1 mM EDTA. In aphid transmission studies, leaf samples were taken from the plants used as transmission sources. Oligonucleotide primers used to amplify the three genomic regions of the SMV isolates were derived from the SMV G2 genomic sequence [26]. P1-coding regions were amplified with primers corresponding to nt 62–88 and complementary to nt 1045–1067. HC/Pro-coding regions were amplified in two segments using primers corresponding to nt 1057–1074 and complementary to nt 1671–1691 and corresponding to nt 1536–1561 and complementary to nt 2425–2450. CP-coding regions were amplified using primers corresponding 8451–8474 and complementary to nt 9355–9378. First-strand cDNA was synthesized by mixing 1.5 μl total extracted nucleic acid, 8.3 μl water and 6.6 pmol of a region specific 3′ primer, and incubating the samples at 97 °C for 5 min, and 40 °C for 10 min. Samples were chilled immediately for 5 min after which 5 μl of 250 mM Tris-HCl (pH 8.3), 375 mM KCl, 15 mM $MgCl_2$, 50 mM DTT, 5 μl 10 mM dNTPs, 2.5 μl 0.1 M DTT, and 100 units moloney murine leukemia virus reverse transcriptase (Promega, Madison, WI) were added and the samples were incubated at 40 °C for 1 h. Distilled water (20 μl) was then added and the samples were heated to 100 °C for 10 min. Amplification mixtures consisted of 5 μl of cDNA, 50 mM KCl, 10 mM Tris-HCl (pH 9.0), 0.1% Triton X-100, 2.5 mM $MgCl_2$, 6 pmol of each primer 300 μM dNTPs and 1 unit *Taq* DNA polymerase in a total volume of 25 μl. The samples were subjected to 30 cycles of 95 °C (45 s), 54 °C (1 min), and 72 °C (1 min) and then stored at 4 °C. PCR products were resolved on 0.8% agarose gels [53]. Infection with SMV was confirmed by the presence of a 929-bp band corresponding to the amplified CP-coding fragment.

RFLP analysis

Nine restriction enzymes, *Alw*I, *Ava*II, *Bfa*I, *Dde*I, *Hae*III, *Nia*III, *Rsa*I, *Sau*96I, and *Taq*I, were used to generate restriction maps of the 929-bp CP DNA fragments that were amplified from 38 SMV strains and isolates. Restriction enzyme digestions were conducted as recommended by the enzyme suppliers in 20 μl and contained 8 to 10 μl of PCR products. The digestions were analyzed on 2% agarose gels.

Nucleotide sequence analysis

The nucleotide sequences of the P1-coding region were determined for 18 isolates, the 5′ half and complete sequence of HC/Pro-coding region was determined for 11 and 14 isolates, respectively. The sequence of the CP-coding region was determined for 25 isolates. Prior to nucleotide sequence analysis, PCR products were purified using PCR purification kit (Qiagen, Valencia, CA). Sequences were determined directly from the PCR products using Prism Big Dye Terminator Cycle Sequencing Ready Reaction Kits (Perkin Elmer, Foster City, CA). All PCR products were sequenced with the primers used to amplify the fragments. In addition, P1 fragments were sequenced with primers corresponding and complementary to nt positions 471–494 and 610–634, respectively; HC/Pro fragments were sequenced with primers corresponding and complementary to nt positions 2031–2051 and 1940–1960, respectively; and CP fragments were sequenced with primers corresponding and complementary to nt positions 8724–8746 and 8925–8945, respectively.

Phylogenetic analysis

Nucleotide sequences were edited using Sequencher 4.0 (Gene Codes, Ann Arbor, MI), translated, and both nucleotide and amino acid sequences were aligned using CLUSTALX [59]. Pair-wise distance and percent identity matrices for each region were calculated using GeneDoc [43]. Phylogenetic relationships between isolates were deduced using the neighbor-joining and parsimony methods in PAUP ver 4.0B8 (Sinauer Associates, Sunderland, MA). Because the topologies of the two sets of trees were similar, only neighbor-joining results are presented. The rates of synonymous and nonsynonymous substitutions were calculated as described by Korber [32] using the SNAP server at http://www.hiv.lanl.gov

Aphid transmission

Aphids, *A. glycines*, were reared in controlled environment chambers on *G. max* var. Williams 82. Alate aphids were collected and starved for at least 30 min before being given 2 min access feeds on leaf tissue infected with eight laboratory SMV isolates G1, G2, G3, G5, G6, G7, G7a, and G7F, and six field SMV isolates 413, 452, 746, 770, 1071, and 1083. The laboratory isolates were maintained by mechanical inoculation, while the field isolates were maintained by transmission with *A. glycines*. For each virus, five aphids were transferred to each of 10 soybean seedlings and allowed to feed for at least 24 h before being fumigated with Vapona (Shell, Houston, TX). Inoculated plants were maintained in a greenhouse for four weeks and assayed for SMV infection by DAS-ELISA. The experiment was conducted three times.

Results

RFLP analysis of the CP region

Amplification of the CP-coding region of 38 SMV isolates produced cDNA fragments of 926 bp. Of the nine restriction enzymes analyzed, five (*Alw*I, *Ava*II, *Hae*III, *Rsa*I, and *Taq*I) produced informative banding patterns (Fig. 1A). Restriction maps were constructed for each enzyme for each isolate and used to generate a phylogram for the 38 isolates (Fig. 1B). This assay also identified four samples that contained more than one virus isolate (e.g., Fig. 1A, lane 3). The phylogenetic analysis of the presence and absence of restriction enzyme sites produced five groups. Group 1 (Fig. 1B) contained only Asian SMV isolates. Groups 4 and 5 contained only North American isolates. Groups 2 and 3 contained both North American and Asian isolates. Based on the groupings from the RFLP analysis, representatives from each group were selected for nucleotide sequence analyses of the P1, HC/Pro and CP-encoding regions.

Sequence and phylogenetic analyses of P1, CP, and HC/Pro-encoding regions

The CP nucleotide sequences were determined from 25 SMV isolates and compared to those of previously published DNA sequences (Table 1). These alignments showed that the CP nucleotide and amino acid sequences of the SMV isolates shared 91 to 99% and 95 to 100% sequence identity, respectively. The levels of sequence identity in the nucleotide and predicted amino acid sequences of the HC/Pro and P1 regions were lower than those for the CP region and separated the SMV isolates into two broad groups according to their geographical origin. The HC/Pro nucleotide sequences of North American SMV isolates were 94 to 98% identical to each other, but shared only 87 to 91% identity with Asian isolates. Among the Asian SMV isolates the HC/Pro nucleotide sequences were 87 to 91% identical. The P1 sequences were the most variable. The P1 nucleotide sequences of North American SMV isolates were 93 to 99% identical to each other, but shared only 84 to 88% identity with Asian isolates. A wide divergence was also seen within the Asian SMV isolates with sequence identities ranging from 85 to 98%.

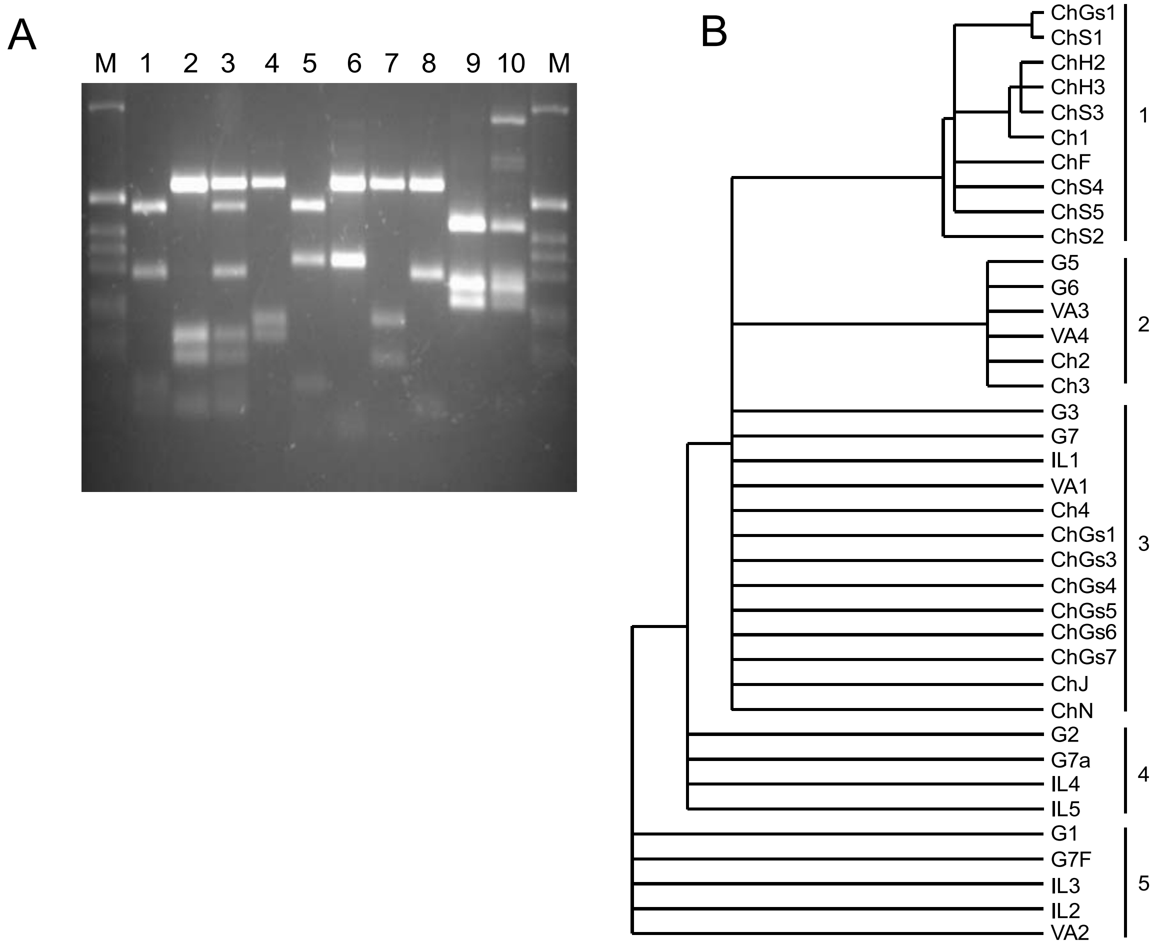

Fig. 1. RFLP analysis of PCR-amplified coat protein regions from 38 SMV isolates. **A** The banding patterns following digestion eight SMV isolates (G1, G3, G4, ChJ, Ch1, ChH2, ChH3, and IL4) with *Rsa*I and two isolates (G3 and G4) with *Dde*I are shown with 1-kd DNA size markers (M). *3* shows the products from a plant infected by two SMV isolates. **B** The presence and absence of restriction sites was used to construct a phylogram using PAUP. The phylogram showed five clusters of isolates (vertical lines and numbers designate groups)

Phylogenetic analyses of the nucleotide sequences showed the same general trends observed in the percent identities of the sequences. Analysis of the entire CP sequences provided little separation of SMV isolates with only one clade with greater than 70% bootstrap support (Fig. 2C). In contrast, trees constructed from the P1 and HC/Pro regions clearly separated North American and Asian isolates into two well supported clades (Fig. 2). In addition, both the P1 and HC/Pro phylograms (Fig. 2A and B) illustrated the relatively small genetic diversity among the North American SMV isolates compared to the larger amount of diversity among Asian SMV isolates.

The ratios of rates of non-synonymous to synonymous substitutions (ω) are indicative of the types and intensities of selection pressures exerted on

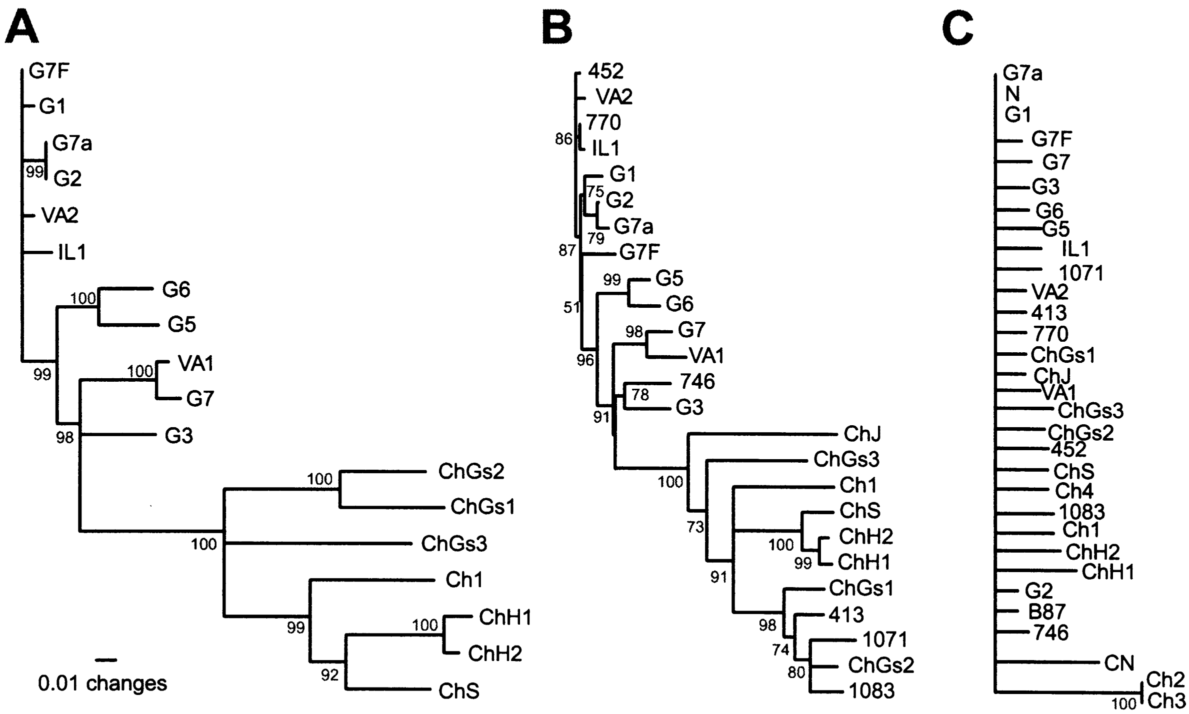

Fig. 2. Phylograms of the nucleotide sequences of the P1 (**A**), HC/Pro (**B**), and CP (**C**) regions of SMV isolates from North America and Asia. The horizontal bar indicates 0.1 nucleotide changes. Bootstrap values greater than 70% based on 1000 replicates are indicated at the corresponding branch points. Branches with bootstrap values less than or equal to 70% were collapsed to polytomies

protein-coding regions. Values of ω greater than 1.0 indicate positive selection; equal to 1.0 neutral selection; and less than 1.0 purifying selection [29]. The average rates of synonymous substitutions were very similar in the three genomic regions analyzed: 0.31, for P1 and 0.28 for HC/Pro and CP. In contrast, ω values differed significantly for each coding region. The average value of ω for the P1, HC/Pro and CP regions were 0.189, 0.027, and 0.012, respectively. The values of ω differed within each of the three proteins (Fig. 3). Within P1, the region corresponding to the conserved serine protease domain [51] (amino acids 240–260) was subject to a high level of purifying selection even though the neighboring regions appeared to be under strong positive selection. With the exception of a region surrounding amino acid 100, all of the HC/Pro had ω values significantly less than 1.0. The first 110 amino acids comprised the cysteine rich region that included the KLSC motif. This region has been shown to be required for aphid transmission, but dispensable for replication and movement of TEV [12, 13]. Of the three regions, the CP was subject to the most intense purifying selection along its length. The amino and carboxyl terminal regions, which corresponded to exposed beta sheet and terminal alpha helical regions of the PVA CP [2], had significantly higher ω values than the internal regions of the protein. The ω values for the P1 (0.22 and 0.19) and CP

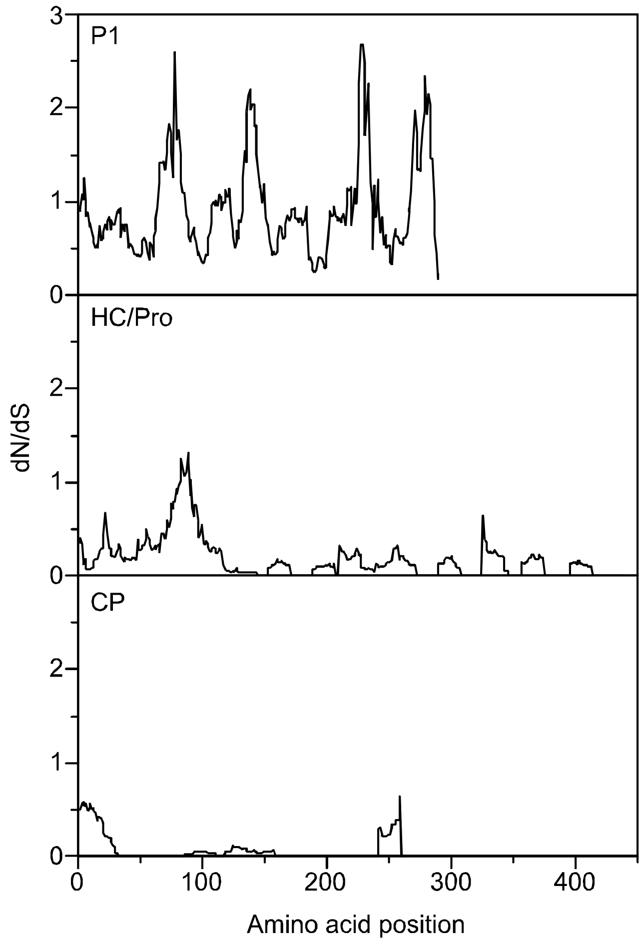

Fig. 3. Plots of ratios of nonsynonymous to synonymous nucleotide substitutions (dN/dS) per in codons of P1, HC/Pro and CP coding sequences. Ratios were calculated using an 18-nt sliding window. The 18 P1 sequences contained 291 codons; the 14 HC/Pro sequences contained 445 codons; and the 31 CP sequences contained 263 codons

(0.009 and 0.011) regions of North American and Asian isolates (respectively) were similar within each region. The complete HC/Pro coding sequence was determined for only three Asian isolates (compared to 11 North American isolates) which did not permit a comparison of the selection pressures with in this region.

Sequences associated with transmission of SMV by Aphis glycines

The aphid transmissibility of 14 SMV isolates was evaluated using *A. glycines* as a vector. Of the eight laboratory isolates analyzed, only one, G3, was transmitted by *A. glycines* in only one of the three transmission tests, but to only one of ten plants (Table 2). In contrast, all six of the field isolates examined were transmitted in all

Table 2. Transmission of SMV isolates by *Aphis Glycines*

SMV Strain	Test 1	Test 2	Test 3	CP DAG	HC/Pro		
					KLSC	PTK	aa subs[a]
G1	0/10	0/10	0/10	edmDA***D***kd	vkKLSCkq	ksPTKrh	1
G2	0/10	0/10	0/10	edmDAX[b]kd	vkKLSCkq	ksPTKrh	0
G3	0/10	0/10	1/10	edmDAGkd	vkKLSCkq	ksPTKrh	2
G5	0/9	0/10	0/10	edmDA***D***kd	vkKLSCkq	ksPTKrh	5
G6	0/10	0/10	0/10	edmDAGkd	vkKLSCkq	ksPTKrh	4
G7	0/10	0/10	0/10	edm***G***A***D***kd	vkKLSCkq	***r***sPTKrh	6
G7a	0/10	0/10	0/10	edmDA***D***kd	vkKL***T***Ckq	ksPTKrh	1
G7F	0/9	0/10	0/10	edm***G***A***D***kd	vkKLSCkq	ksPTKrh	5
413	9/10	8/10	3/10	edmDAGkd	vkKLSCkq	ksPTKrh	NA[d]
452	7/10	6/10	4/10	edmDAGkd	vkKLSCkq	ksPTKrh	NA
746	4/10	2/10	3/10	edmDAGkd	vkKLSCkq	ksPTKrh	NA
770	ND	ND	ND	edmDAGkd	vkKLSCkq	ksPTKrh	NA
1071	7/10	8/10	6/10	edmDAGkd	vkKLSCkq	ksPTKrh	NA
1083	6/10	10/10	7/10	edmDAGkd[e]	vkKLSCkq	ksPTKrh	NA

[a]Number of amino acid substitutions not present in at least one of the transmissible isolates

[b]A unique nucleotide sequence could not be obtained from this isolate at this position. Both DAG and DAD triplets were predicted from sequence

[c]Not determined

[d]Not applicable

[e]All other seed-borne and field isolates had amino acid sequences identical to those of the efficiently transmitted isolates listed

three transmission tests with efficiencies ranging from 20 to 100%. The predicted amino acid sequence motifs of HP/Pro and CP that have been associated with aphid transmission of other potyviruses [46], i.e., KLSC (amino acids 53–56; KITC in many other potyviruses) and PTK (amino acids 309–310) in HC/Pro and DAG (amino acids 10–12) in CP, were conserved and identical in all field isolates (Table 2). With the exception of a D to V amino acid substitution in SMV-G7 at position 101, all amino acid sequence changes in the CP sequences were within the first 27 amino acids, which included the conserved DAG motif. Laboratory isolates G1, G2, G5, G7, and G7F had mutations in the DAG motif. Isolate G7a had mutations in both the DAG and KLSC motifs. In addition to the aphid nontransmissible strain SMV N, which was previously shown to lack a DAG triplet [15], three other previously reported SMV CP sequences, two from China (X63771 and U25673) and one from Japan (D88616), also lacked DAG motifs.

The HC/Pro PTK motif was conserved in all of the sequences analyzed. Isolates G3 and G6 retained all three conserved motifs, but were transmitted poorly. Neither G3 nor G6 had amino acid substitutions in the CP that differentiated them from the field isolates. Both G3 and G6 had amino acid substitutions outside of the conserved HC/Pro motifs that were not present in the sequences of any of the efficiently transmitted isolates (Table 2). The predicted amino acid sequence of

SMV G3 HC/Pro had substitutions of A to T and F to L at positions 110 and 206, respectively. The predicted amino acid sequence of SMV G6 HC/Pro had substitutions of N to K, P to S, T to P, and R to Q at amino acid positions 34, 241, 271, and 374, respectively.

Discussion

In this study, we examined the variability in nucleotide sequence and aphid transmissibility of North American and Asian SMV isolates. Because of the number of viruses analyzed and the potential difficulty of correlating symptom severity with specific nucleotide sequence variation, no attempt was made to associate sequences with disease phenotypes. For example, symptom phenotypes of *Tobacco mosaic virus* [30] and *Cucumber mosaic virus* [57] are dramatically altered by single amino acid substitutions, which indicate that symptoms are not always a reliable indicator of genetic relatedness. Still, information about the severity of the symptoms produced by virus strains on a range of hosts is valuable for defining their potential agronomic impact.

As has been reported for other species within the family *Potyviridae*, neither RFLP nor nucleotide sequence analyses of RT-PCR amplified CP-coding regions provided good indications of the genetic relatedness of the SMV isolates examined. However, RFLP analysis did identify mixed isolates, which could not be detected using the differential set of soybean indicator lines. In contrast, analysis of sequences from P1 and HC/Pro-coding regions clearly showed differences in the genetic variability of North American and Asian SMV isolates. Similar levels of variability in P1 regions have been noted for several other members of the family *Potyviridae* including, *Lettuce mosaic virus* [50], *Maize dwarf mosaic virus* [31], *Papaya ringspot virus* [63], *Potato virus A* (PVA) [27], *Potato virus Y* (PVY) [60], *Sweet potato feathery mottle virus* [52], *Yam mosaic virus* [1, 16], and *Zucchini yellow mosaic virus* [34]. Even though several activities have been associated with the P1 protein (e.g., RNA binding and protease), its role in the virus replication cycle remains to be established. The higher levels of variability tolerated in the P1 compared to the CP-coding regions suggests either that there are fewer structural constraints on the protein or that it interacts with a varying set of host factors.

The North American SMV isolates analyzed in this study were much less variable than those from Asia, where cultivated soybeans originated [22]. Similarly, Xu et al. [67] reported that Chinese SMV isolates showed greater diversity in symptom and disease severity than North American isolates. The difference in diversity between the two groups of isolates is likely related to differences in the balance between purifying and positive selection pressures in the two environments and founder effects that resulted from the introduction of relatively few SMV variants into North America. Similar founder effects have been studied in great detail in Human immunodeficiency virus [39] and cited as the probable cause of low diversity in other plant virus populations [54]. However, since soybeans were first introduced into the U.S. in 1789 [23], there have been several importations of soybean germplasm that could have brought additional SMV diversity. The high error rate of replication of viral RNAs [14] and the widespread cultivation of soybeans in North America would seem to have provided opportunities for sequence diversity to accumulate. Hall et al. [17] estimated that *Wheat streak mosaic virus* (WSMV), another member of the family *Potyviridae*, accumulates mutations at a rate of about 0.36 nt changes/genome/passage. In an analysis of the diversity of North American WSMV isolates, Stenger et al. [58] calculated that the current level of diversity present in WSMV in North America (up to 21%) could be explained by the 100 year history of wheat monoculture in North America. Soybean has a much shorter history of widespread cultivation in North America, which may explain its lower (less than 10%) variability in the CP region.

Analysis of the mutation rates of viruses in different hosts has shown that the size of the quasispecies clouds for plant viruses can be host dependent [55]. Hence, the availability and diversity of primary and alternative hosts may influence the variability of North American SMVs. In North America, the principle host of SMV is cultivated soybean, which have much lower genetic diversity than soybeans grown in Asia [35]. In addition, Asia is home to *G. soja*, the progenitor of cultivated soybean, which is a host to SMV, widely distributed throughout Asia, and much more diverse than cultivated soybean [40, 66]. Even so, the impact of host diversity on the diversity of potyvirus genomes has not been established. For example, Moury et al. [42] did not detect an effect of host range on the diversity of CP-coding regions of seven other members of the family *Potyviridae*.

Even though the rates of synonymous nucleotide substitutions were very similar in the three regions of the SMV genome analyzed, the ratios of the rates of nonsynonymous to synonymous substitutions indicated that there was much stronger purifying selection in the CP-encoding region than in the other two genomic regions. Hence, vector transmission and the requirement to assemble stable virions impose intense purifying selection pressures on CP sequences. The effects of the selection of aphid transmissibility are illustrated by the data presented in Table 2, which showed that the North American field and more variable Asian seedborne SMV isolates had identical sequences in regions surrounding the amino acid motifs associated with aphid transmission. In contrast, the less divergent North American laboratory isolates that had been transmitted mechanically contained mutations in the DAG and KLSC motifs. In a similar study of 13 PVA isolates, Rajamäki et al. [49] identified five PVA isolates that were not transmitted by *Myzus percicae* all of which had mutations in the DAG motif. Unlike the SMV isolates analyzed here, the PVY HC/Pro region had the lowest ω value of the P1, HC/Pro and CP regions. The ω value reported for the PVY P1 region (0.23) [42] was very similar to that for SMV isolates (0.19), but ω for the PVY CP was much higher (0.162) than that SMV CP (0.012) possibly indicating that the CPs of the two viruses are subject to different intensities of purifying selection pressure.

Vertical transmission through seed imposes a different set of selection pressures than transmission by aphids. Viruses transmitted through seed are dependent upon infected hosts to produce viable offspring, which gives viruses with low or moderate virulence a selective advantage [5]. While aphid transmission selects viruses that have retained tertiary structures within HC/Pro and CP that are required for aphid transmissibility [47], it also facilitates the survival and dispersal of more diverse and potentially more severe virus isolates. As *A. glycines* extends its range in North America, the incidence and potentially virulence of horizontally transmitted SMV isolates could increase. In conclusion, the differences in diversity of North American and Asian isolates likely results from the interplay between many factors, including cultural practices, founder effects, prevalence of alternative hosts, and modes of transmission.

Acknowledgements

This work was supported by the United States Department of Agriculture–Agricultural Research Service and the Illinois Council for Food and Agricultural Research. Mention of a trademark, proprietary product, or vendor does not constitute a guarantee or warranty of the product by the USDA or the University of Illinois and does not imply its approval to the exclusion of other products or vendors that may also be suitable.

References

1. Aleman-Verdaguer ME, Goudou-Urbino C, Dubern J, Beachy RN, Fauquet C (1997) Analysis of the sequence diversity of the P1, HC, P3, NIb and CP genomic regions of several yam mosaic potyvirus isolates: implications for the intraspecies molecular diversity of potyviruses. J Gen Virol 78: 1253–1264
2. Baratova LA, Efimov AV, Dobrov EN, Fedorova NV, Hunt R, Badun GA, Ksenofontov AL, Torrance L, Jarvekulg L (2001) *In situ* spatial organization of *Potato virus A* coat protein subunits as assessed by tritium bombardment. J Virol 75: 9696–9702
3. Berger PH, Wyatt SD, Shiel PJ, Silbernagel MJ, Druffel K, Mink GI (1997) Phylogenetic analysis of the *Potyviridae* with emphasis on legume-infecting potyviruses. Arch Virol 142: 1979–1999
4. Berger PH, Barnettt OW, Brunt AA, Colinet D, Edwardson JR, Hammond J, Hill JH, Jordan RL, Kashiwazaki S, Makkouk K, Morales G, Rybicki EP, Spence N, Ohki ST, Uyeda I, van Zaayen A, Vetten HJ (2000) Family *Potyviridae*. In: van Regenmortel HHV, Fauquet CM, Bishop DHL, Carstens EB, MEstes MK, Lemon SM, Maniloff J, Mayo MA, McGeoch DJ, Pringle CR, Wickner RB (eds) Virus Taxonomy: Seventh Report of the International Committee on Taxonomy of Viruses. Academic Press, New York, pp 703–724
5. Bergstrom CT, McElhany P, Real LA (1999) Transmission bottlenecks as determinants of virulence in rapidly evolving pathogens. Proc Natl Acad Sci USA 96: 5095–5100
6. Blanc S, Ammar ED, Garcia-Lampasona S, Dolja VV, Llave C, Baker J, Pirone TP (1998) Mutations in the potyvirus helper component protein: effects on interactions with virions and aphid stylets. J Gen Virol 79: 3119–3122
7. Buzzell RI, Tu JC (1984) Inheritance of soybean resistance to soybean mosaic virus. J Heredity 75: 82
8. Cho EK, Goodman RM (1979) Strains of soybean mosaic virus: classification based on virulence in resistant soybean cultivars. Phytopathology 69: 467–470

9. Cho EK, Goodman RM (1982) Evaluation of resistance in soybeans to soybean mosaic virus strains Cultivars. Crop Sci 22: 1133–1136
10. Chu R, Leng X, Bao Y, Pu Z, Pan N, Chen Z (1992) Amplification of soybean mosaic virus coat protein gene by polymerase chain reaction and its sequence analysis. Acta Bot Sin 34: 523–528
11. Clark AJ, Perry KL (2002) Transmissibility of field isolates of soybean viruses by *Aphis glycines*. Plant Dis 86: 1219–1222
12. Dolja VV, Herndon KL, Pirone TP, Carrington JC (1993) Spontaneous mutagenesis of a plant potyvirus genome after insertion of a foreign gene. J Virol 67: 5968–5975
13. Dolja VV, Hong J, Keller KE, Martin RR, Peremyslov VV (1997) Suppression of potyvirus infection by coexpressed closterovirus protein. Virology 234: 243–252
14. Domingo E, Holland JJ (1997) RNA virus mutations and fitness for survival. Annu Rev Microbiol 51: 151–178
15. Eggenberger AL, Stark DM, Beachy RN (1989) The nucleotide sequence of a soybean mosaic virus coat protein-coding region and its expression in *Escherichia coli*, *Agrobacterium tumefaciens* and tobacco callus. J Gen Virol 70: 1853–1860
16. Fuji S, Nakamae H (2000) Complete nucleotide sequence of the genomic RNA of a mild strain of Japanese yam mosaic potyvirus. Arch Virol 145: 635–640
17. Hall JS, French R, Morris TJ, Stenger DC (2001) Structure and temporal dynamics of populations within wheat streak mosaic virus isolates. J Virol 75: 10231–10243
18. Hill JH, Benner HI, Permar TA, Bailey TB, Andrews RE Jr, Durand DP, Van Deusen RA (1989) Differentiation of soybean mosaic virus isolates by one-dimensional trypsin peptide maps immunoblotted with monoclonal antibodies. Phytopathology 79: 1261–1265
19. Hill JH, Benner HI, Vandeusen RA (1994) Rapid differentiation of soybean mosaic virus isolates by antigenic signature analysis. J Phytopathol 142: 152–162
20. Hill JH (1999) Soybean Mosaic Virus. In: Hartman GL, Sinclair JB, Rupe JC (eds) Compendium of soybean diseases. The American Phytopathological Society, St. Paul, MN, pp 70–71
21. Hill JH, Alleman R, Hogg DB, Grau CR (2001) First report of transmission of *Soybean mosaic virus* and *Alfalfa mosaic virus* by *Aphis glycines* in the new world. Plant Dis 85: 561
22. Hymowitz T (1970) On the domestication of soybean. Econ Bot 24: 408–421
23. Hymowitz T, Harlan JR (1983) Introduction of soybean to North America by Samuel Bowen in 1765. Econ Bot 37: 371–379
24. Jain RK, McKern NM, Tolin SA, Hill JH, Barnett OW, Tosic M, Ford RE, Beachy RN, Yu MH, Ward CW (1992) Confirmation that fourteen potyvirus isolates from soybean are strains of one virus by comparing coat protein peptide profiles. Phytopathology 82: 294–299
25. Jayaram C, Hill JH, Miller WA (1991) Nucleotide sequences of the coat protein genes of two aphid-transmissible strains of soybean mosaic virus. J Gen Virol 72: 1001–1003
26. Jayaram C, Hill JH, Miller WA (1992) Complete nucleotide sequences of two soybean mosaic virus strains differentiated by response of soybean containing the *Rsv* resistance gene. J Gen Virol 73: 2067–2077
27. Kekarainen T, Merits A, Oruetxebarria I, Rajamaki M, Valkonen JP (1999) Comparison of the complete sequences of five different isolates of *Potato virus A* (PVA), genus *Potyvirus*. Arch Virol 144: 2355–2366
28. Kihl RAA, Hartwig EE (1979) Inheritance of reaction to soybean mosaic virus in soybeans. Crop Sci 19: 372–375

29. Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge
30. Knorr DN, Dawson WO (1988) A point mutation in the tobacco mosaic virus capsid protein gene induces hyposensitivity in *Nicotiana sylvestris*. Proc Natl Acad Sci USA 85: 170–174
31. Kong P, Steinbiss HH (1998) Complete nucleotide sequence and analysis of the putative polyprotein of maize dwarf mosaic virus genomic RNA (Bulgarian isolate). Arch Virol 143: 1791–1799
32. Korber B (2000) HIV signature and sequence variation analysis. In: Rodrigo AG, Learn GH (eds) Computational analysis of HIV molecular sequences. Kluwer Academic Publishers, Dordrecht, pp 55–72
33. Lee KC, Mahtani PH, Chng CG, Wong SM (1997) Sequence and phylogenetic analysis of the cytoplasmic inclusion protein gene of zucchini yellow mosaic potyvirus – its role in classification of the *Potyviridae*. Virus Genes 14: 41–53
34. Lee KC, Wong SM (1998) Variability of P1 protein of zucchini yellow mosaic virus for strain differentiation and phylogenetic analysis with other potyviruses. DNA Seq 9: 275–293
35. Li Z, Qiu L, Thompson JA, Welsh MM, Nelson RL (2001) Molecular genetic analysis of U.S. and Chinese soybean ancestral lines. Crop Sci 41: 1330–1336
36. Lim SM (1985) Resistance to soybean mosaic virus in soybeans. Phytopathology 75: 199–201
37. Liu J, Peng X, Li L, Mang K (1993) Cloning of coat protein gene of soybean mosaic virus and its expression in *Escherichia coli*. Chin J Biotechnol 9: 143–149
38. Lucas BS, Hill JH (1980) Characteristics of the transmission of three soybean mosaic virus isolates by *Myzus persicae* and *Rhopalosiphum maidis*. J Phytopathol 99: 47–53
39. Lukashov VV, Karamov EV, Eremin VF, Titov LP, Goudsmit J (1998) Extreme founder effect in an HIV type 1 subtype A epidemic among drug users in Svetlogorsk, Belarus. AIDS Res Hum Retroviruses 14: 1299–1303
40. Maughan PJ, Maroof MAS, Buss GR, Huestis GM (1996) Amplified fragment length polymorphism (AFLP) in soybean – species diversity, inheritance, and near-isogenic line analysis. Theor Appl Genet 93: 392–401
41. McNeil JE, French R, Hein GL, Baenziger PS, Eskridge KM (1996) Characterization of genetic variability among natural populations of wheat streak mosaic virus. Phytopathology 86: 1222–1227
42. Moury B, Morel C, Johansen E, Jacquemond M (2002) Evidence for diversifying selection in Potato virus Y and in the coat protein of other potyviruses. J Gen Virol 83: 2563–2573
43. Nicholas HB Jr, Deerfield DWI (1997) GeneDoc: analysis and visualization of genetic variation. EMBNEWNEWS 4: 14
44. Omunyin ME, Hill JH, Miller WA (1996) Use of unique RNA sequence-specific oligonucleotide primers for RT-PCR to detect and differentiate soybean mosaic virus strains. Plant Dis 80: 1170–1174
45. Peng YH, Kadoury D, Gal-On A, Huet H, Wang Y, Raccah B (1998) Mutations in the HC-Pro gene of zucchini yellow mosaic potyvirus: effects on aphid transmission and binding to purified virions. J Gen Virol 79: 897–904
46. Pirone TP, Perry KL (2002) Aphids: Non-persistent transmission. In: Plumb RT, Callow JA (eds) Advances in botanical research: plant virus vector interactions, vol 36. Academic Press, London, pp 1–19
47. Power AG (2000) Insect transmission of plant viruses: a constraint on virus variability. Curr Opin Plant Biol 3: 336–340

48. Raccah B, Blanc S, Huet H (2001) Molecular basis of vector transmission: Potyvirus. In: Harris K, Duffus JE, Smith OP (eds) Virus-insect-plant interactions. Academic Press, San Diego
49. Rajamäki M, Merits A, Rabenstein F, Andrejeva J, Paulin L, Kekarainen T, Kreuze JF, Forster RLS, Valkonen JPT (1998) Biological, serological, and molecular differences among isolates of potato a potyvirus. Phytopathology 88: 311–321
50. Revers F, Yang SJ, Walter J, Souche S, Lot H, Le Gall O, Candresse T, Dunez J (1997) Comparison of the complete nucleotide sequences of two isolates of lettuce mosaic virus differing in their biological properties. Virus Res 47: 167–177
51. Ryan MD, Flint M (1997) Virus-encoded proteinases of the picornavirus super-group. J Gen Virol 78: 699–723
52. Sakai J, Mori M, Morishita T, Tanaka M, Hanada K, Usugi T, Nishiguchi M (1997) Complete nucleotide sequence and genome organization of sweet potato feathery mottle virus (S strain) genomic RNA: the large coding region of the P1 gene. Arch Virol 142: 1553–1562
53. Sambrook J, Russell DW (2001) Molecular Cloning: A Laboratory Manual, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
54. Sanchez Campos S, Diaz JA, Monci F, Bejarano ER, Reina J, Navas-Castillo J, Aranda MA, Moriones E (2002) High genetic stability of the begomovirus *Tomato yellow leaf curl Sardinia virus* in southern Spain over an 8-year period. Phytopathology 92: 842–849
55. Schneider WL, Roossinck MJ (2001) Genetic diversity in RNA virus quasispecies is controlled by host-virus interactions. J Virol 75: 6566–6571
56. Shi XM, Miller H, Verchot J, Carrington JC, Vance VB (1997) Mutations in the region encoding the central domain of helper component-proteinase (HC-Pro) eliminate potato virus X/potyviral synergism. Virology 231: 35–42
57. Shintaku MH, Zhang L, Palukaitis P (1992) A single amino acid substitution in the coat protein of cucumber mosaic virus induces chlorosis in tobacco. Plant Cell 4: 751–757
58. Stenger DC, Seifers DL, French R (2002) Patterns of polymorphism in *Wheat streak mosaic virus*: sequence space explored by a clade of closely related viral genotypes rivals that between the most divergent strains. Virology 302: 58–70
59. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25: 4876–4882
60. Tordo VM, Chachulska AM, Fakhfakh H, Le Romancer M, Robaglia C, Astier-Manifacier S (1995) Sequence polymorphism in the 5′NTR and in the P1 coding region of potato virus Y genomic RNA. J Gen Virol 76: 939–949
61. Verchot J, Koonin EV, Carrington JC (1991) The 35-kDa protein from the N-terminus of the potyviral polyprotein functions as a third virus-encoded proteinase. Virology 185: 527–535
62. Wang CH, Bau HJ, Yeh SD (1994) Comparison of the nuclear inclusion b protein and coat protein genes of five papaya ringspot virus strains distinct in geographic origin and pathogenicity. Phytopathology 84: 1205–1210
63. Wang CH, Yeh SD (1997) Divergence and conservation of the genomic RNAs of Taiwan and Hawaii strains of papaya ringspot potyvirus. Arch Virol 142: 271–285
64. Wang RY, Ghabrial SA (2002) Effect of aphid behavior on efficiency of transmission of *Soybean mosaic virus* by the soybean-colonizing aphid, *Aphis glycines*. Plant Dis 86: 1260–1264
65. Ward CW, McKern NM, Frenkel MJ, Shukla DD (1992) Sequence data as the major criterion for potyvirus classification. Arch Virol [Suppl 5]: 283–297

66. Xu B (1989) Analysis of ecotypes and their geographical distribution of wild soybean (*G. soja*) in the world. Ecological Acta 13: 114–120
67. Xu Z, Polston JE, Goodman RM (1986) Identification of soybean mosaic, southern bean mosaic and tobacco ringspot viruses from soybean in the People's Republic of China. Ann Appl Biol 108: 51–57

Author's address: Leslie L. Domier, United States Department of Agriculture–Agricultural Research Service, 1102 S. Goodwin Ave., University of Illinois at Urbana/Champaign, Urbana, IL 61801, U.S.A.; e-mail: l-domier@uiuc.edu